Institution: University of Huddersfield	
Unit of Assessment: 12 (Engineering)	
Title of case study: Improving Rail Safety by Enabling the Adoption of Modern Techniques by the Industry	
Period when the underpinning research was undertaken: 2014-2020	
Details of staff conducting the underpinning research from the submitting unit:	

Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Coen van Gulijk	Professor	2014–2019 ; V Prof. 2020
Peter Hughes	Principal Researcher	2014- present
Miquel Figueres	Research Assistant	2014–2019 ; associate 2020
Rawia El Rashidy	Research Assistant	2014- present
Julian Stow	Director	2012- present

Period when the claimed impact occurred: 2014-2020

Is this case study continued from a case study submitted in 2014? No

1. Summary of the impact

Safety is a major concern on railway networks. Traditionally operators had more data available to them than they were able to analyse and interpret, and this led to risks being left unattended, putting passengers and railway staff potentially at risk of injuries and fatality.

The impact of research by the University of Huddersfield (UoH) has been a significant improvement in the effectiveness of railway safety management systems, allowing new safety insights to be extracted from unstructured data, new visualization techniques to be adopted and new tools to be developed that provide rapid analysis of operational data to identify changes in safety risk.

Railway undertakings in the UK were able to improve their safety management (including Network Rail and LNER), benefitting passengers and policy makers (Rail Safety and Standards Board in the UK and the European Railway Agency), plus rail operators in mainland Europe (e.g. RENFE in Spain).

2. Underpinning research

Railways rely on safety management systems to capture data (both structured and unstructured), which is then used by safety experts to better understand and model railway system safety risk. This understanding enables the development of strategies to improve safety, thus reducing accidents, injuries and fatalities.

In 2012 the Rail Safety and Standards Board (RSSB) identified that the volume of data produced by train operators meant that it could not all be analysed using “traditional” techniques and this led to missed opportunities to improve safety. The RSSB Rail Technology Strategy (2012), required the industry to extract more value from the data it had available. University of Huddersfield (UoH) research identified how to apply emerging digital data analysis techniques, to move industry practices from risk-based strategies to ones based on prediction.

The research was carried out in the Institute of Railway Research (IRR), based at UoH. It was performed by Dr Coen Van Gulijk (Professor of Railway Safety and Risk at UoH since 2014), Peter Hughes (Principal Researcher since 2014), Miguel Figueres and Rawia El Rashidy (Research Assistants since 2014) and Julian Stow (Director IRR at UoH since 2012).

The research focused on three key areas:

Text Analysis: As part of a strategy to identify “hidden” risks, railway employees across the UK produce over 300,000 near-miss reports each year. A near-miss report is a free-text report containing details of something that could have caused an accident. Processing this volume of text-based records by hand required a large team of individuals to extract the information that led to safety lessons.

Research by Figueres [R1] and Hughes [R2,3] (2018) showed different approaches of automated text analysis, referred to as Natural Language Processing (NLP), could be a used for the analysis of safety reports. Standard NLP techniques do not perform well with railway jargon or poor spelling (both commonplace in safety reports). A novel NLP approach was developed that dramatically improved the outputs from NLP on this unstructured data. The solution utilized naïve ontology word nets to manually create “clouds” of safety-related words. These were used to analyse the near-miss reports and the word clouds were refined so that individual risk types could be extracted. The team used a state-of-the-art Graph database, which was infinitely scalable and able to process thousands of documents in a matter of seconds. The database also had a graphical user interface that displayed the results of the analyses as bubble diagrams, which were easy for non-experts to interpret. The move from manual to automated analysis reduced the processing time from days, or even weeks, to hours.

Train and Signal Network Data: A train that fails to react to a red signal is known as a Red Signage Passage (also Signal Passed at Danger, or SPAD) and is the most dangerous safety incident on the railways – when a high-speed train SPADs it can collide with other trains, potentially killing hundreds of people. Risk modelling for SPADs traditionally used statistical methods to analyse exposure to SPAD risk, but this is, by definition, imprecise.

The research [R4] (2016) explored how exposure to SPAD risk could be identified from data already collected by train recorders and other infrastructure. One source of useful data was Network Rail’s (NR) Train Describer data-feed, gathered at signals, to detect trains that came dangerously close to a SPAD. An algorithm was developed to use the data to report when a train failed (or nearly failed) to stop at a red signal. The research findings showed, for the first time, the prevalence of dangerous approaches to red signals across the UK network at the individual signal level. The detection system covered the whole of the digitally enabled rail network, amounting to about 60% of NR’s tracks. A practical, safety-related application was developed for RSSB, and statistical guesswork on the exposure to SPAD risk was much reduced.

In 2017 the research was extended by combining on-train data with the signal data feed [R5] in a transparent Graph database. In 2019, new data from the West Coast mainline was added, in collaboration with Virgin Trains. The data was manually extracted from the on-board recorder at the end of a locomotive’s shift and taken to Huddersfield to be extracted, modified and interpreted. This extra locomotive-specific data meant it was possible to take the true speed of travel into account. Since the braking behaviour of a train differs depending on whether the track is dry or wet (due to rain or leaves), slippery tracks could be detected and near-misses could be identified [R6].

Integration of Data into a Visualization Tool: Risk management is a key task for railway operators. Individual companies employ dedicated teams to deal with risk monitoring, incident investigation, assessing protection measures and reporting. The BowTie was a risk management tool used by the Chemical Industry to visualize risk by creating a graphical representation of the “risk space”. At the start of this study the railway industry was beginning to adopt this as a manual tool, with the BowTie drawn on paper by the analyst.

The transparent database approach [R5] (2014) made it possible to integrate the different railway-specific data-sources, e.g. text analysis and the train and signal network data, into commercial software by using an interface built by the UoH team. This enabled a BowTie risk model to be automatically created and to be updated as new data arrived.

3. References to the research

The papers below introduce data techniques for non-engineering safety solutions in the safety science domain. They are among the very first ones to present practical techniques for Natural Language Programming, and data techniques for risk-scenario detection. The work was published in leading international journals for safety in general and safety in the railways.

1. Figueres-Esteban M, Hughes P and Van Gulijk C (2016) Visual analytics for text-based railway incident reports. Safety Science **89**: 72–76. https://doi.org/10.1016/j.ssci.2016.05.009

2. Hughes P, Figueres-Esteban M and Van Gulijk C (2017) From free-text to structured safety management: Introduction of a semi-automated classification method of railway hazard reports to elements on a bow-tie diagram. Safety Science **110**: 11–19. https://doi.org/10.1016/j.ssci.2018.03.011

3. Hughes P, Robinson R, Figueres-Esteban M and Van Gulijk C (2019) Extracting safety information from multi-lingual accident reports using an ontology-based approach. Safety Science **118**: 228–297. https://doi.org/10.1016/j.ssci.2019.05.029

4. Zhao Y, Stow J and Harrison C (2016) Estimating the frequency of trains approaching red signals: a case study for improving the understanding of SPAD risk, IET Intelligent infrastructures **10-9**: 579–586. https://doi.org/10.1049/iet-its.2015.0052

5. El Rashidy R, Hughes P, Figueres-Esteban M, Harrison C and Van Gulijk C (2017) A big data modeling approach with graph databases for SPAD risk. Safety Science **110**: 75–79. https://doi.org/10.1016/j.ssci.2017.11.019

6. El Rashidy R, Hughes P & Van Gulijk, C (2019) Detection of high-speed red-aspect approaches using multi-data approach. Safety Science **120**: 583–588. https://doi.org/10.1016/j.ssci.2019.07.027

4. Details of the impact

The research led to impact on rail operators and their customers in the UK and across Europe, by improving safety (and punctuality) across the networks. European railway policy was also influenced.

The impacts can be summarized under four headings:

1. Reduction in missed red signals
2. Extracting safety lessons from close calls
3. Preparing for digital risk management
4. Policy influence

Reduction in Missed Red Signals

The research on modelling the prevalence of SPADs [R4,5] was game-changing. For the first time, it was demonstrated that it is possible to replace statistical models, with measurements of actual train behaviour across the UK network. The Rail Safety and Standards Board (RSSB) used the Institute of Railway Research (UoH-IRR) algorithms to develop its own tool to exploit the data. The Risk Aspect Approaches to Signals (RAATS) tool was published through a web service in 2019 [E6], thus making it possible for anyone working on the railways to benefit from superior insights into "a significant safety risk to the railway" [E6]. The web service has had 200 registered users since it was launched, which is the equivalent of about three registered users from each of the over 60 UK train operating companies.

Railway operators used the tool in two ways. Planners used it to identify pinch points on the network (where a high volume of trains approach a signal when it is red) and then planned these situations out of the timetable. This greatly reduced SPAD risk and improved the overall flow of traffic on the network, thus increasing punctuality [E7]. The punctuality improvement is significant because each “delay minute” costs a rail operator up to £3,000. Safety practitioners used the tool to analyse data collected on the frequency of SPADs, to understand the underlying probability of a potentially dangerous red signal approach and whether this was a contributory factor to the SPAD event.

The addition of new, direct from trains, data to the model [R6] isolated two types of red aspect approaches that are particularly dangerous (i) green-light gambling (assuming that the signal will clear in time) and (ii) slippery conditions. Such data is now used by risk analysts at RSSB, Network Rail and other GB railway operators. RSSB stated that “without the research from the IRR this advancement would not have been possible”' and that “the benefits in reduction in safety risk, both now and in the future, [..] would not have been realised” [E7] (2020).

Extracting Safety Lessons from Close Calls

The structured investigation [R1] and automated classification of close-call data using Natural Language Processing (NLP) [R2,3] has been used in non-standard risk assessments by RSSB. It reported that the data helped to identify risks that were undervalued by the traditional “triage” process. For example, the RSSB’s Professional Head of Safety and Intelligence highlighted its role in identifying, for the first time, risks in relation to vegetation management on slopes [E7]. The method was also used for monitoring purposes [R2] to classify how safety risks increased or decreased over time. RSSB has used the tool to analyse over a million records, stating it “enables rapid interrogation of the data” and “delivers structured results that can more readily support robust and actionable changes to safety practices” [E7]. Similar work was continued by RSSB, as demonstrated in its 2019 report T1152 [E8].

The chair of the steering committee for the strategic partnership that oversaw the research, worked for Network Rail (NR), and encouraged NLP development because of the improvements in speed and accuracy of risk identification that the techniques delivered. The IRR developed a bespoke text-analysis tool for NR, called C-CAT. It used text analysis to classify 300,000 close-call records against 26 risk categories. NR had a team of 30 full-time employees analysing the reports and by removing the need for humans to read every single report, the analysis time was greatly reduced. Importantly C-CAT also helped re-classify the 30% of entries that were unintentionally reported in the “general” category [E9], hence automating the correction of 300,000 entries (2018).

In Spain, the national railway operator, RENFE, reported, “we are investing [..] EUR 0.25m in the development of a ‘close call’ reporting system that utilises the ontology and data processing techniques based on the IRR research” [E5].

Preparing for digital risk management using the BowTie tool

The research findings on the BowTie tool [R5] demonstrated ways to populate commercial risk BowTies with consistent railway-specific data, gathered from disparate automated sources.

LNER worked with IRR to upgrade their safety management system. They are now recognized as having the leading BowTie experts on the GB railways and this is evidenced by their bow-tie manual [E10], which is used across the sector. LNER’s use of advanced digitally enabled BowTie software for its safety management system has provided it with unprecedented oversight of operational risks.

The French railway company SNCF started a digital programme in France in 2016. Adding their own digital developments to the IRR foundation, they initiated a national safety programme called “Prisme”. The Head of SNCF Railway System Safety stated, “The research from IRR and the development of the bowtie tool allow SNCF to better identify where to invest funds to improve, or maintain, safety of operations". He added, "Without the research from IRR, this improvement would not have happened!" [E4].

In 2019, RENFE in Spain started building a new safety system based on BowTies. Their Head of Safety said, "IRR research on the use of bowtie analyses in safety management systems has been adopted by RENFE and we are now working [..] on a EUR 1M two-year project to connect our operational processes and procedures with safety controls defined within a bowtie analysis" [E5].

IRR researchers spoke at a number of industry-specific BowTie events, which led to an increase in sales [E10] for the organizations that licensed the technology, such as LNER, RSSB and CGE-Risk (a provider of BowTie software).

Policy Influence
The European Railway Agency (ERA) participated at the BDRA (Big Data Risk Analysis) symposia in Huddersfield (2016 and 2017) [E1] and invited the University of Huddersfield to be the only academic partner in their four workshops on the digitization of railways (2016–18). ERA acknowledged the input from UoH workshops on their policy, in their reports [E2, E3].

5. Sources to corroborate the impact

[E1]: Editorial in SaRS journal that printed the proceedings of the 1st industrial BDRA conference tandfonline.com/doi/full/10.1080/09617353.2016.1252082
[E2]: Slide by ERA identifying the inputs for the CSM-ASLP project reviewing the European incident registration system for ERA.
[E3]: Meeting minutes for 3rd industry workshop on DSD demonstrating industry engagement and impact into the ERA.
[E4]: Testimonial by SNCF supporting our claim for the impact of our work on SNCF (France). https://www.era.europa.eu/sites/default/files/events-news/docs/sncf_frederic_delorme_en.pdf
[E5]: Testimonial by RENFE supporting our claim for the impact of our work on RENFE (Spain).
[E6]: the online RAATS tool, accessible through: https://www.rssb.co.uk/en/safety-and-health/improving-safety-health-and-wellbeing/Rail-Risk-Toolkit/Red-Aspect-Approaches-to-Signals-Toolkit
[E7]: Testimonial by RSSB supporting our claim for improved management of railway safety risk in the UK.
[E8]: The cover and executive summary of T1152 report demonstrating that RSSB continued research that was originally designed in this work
[E9]: Delivery note and manual for C-Cat delivered to Network Rail.
[E10]: LNER BowTie risk assessment method.